

# (12) UK Patent Application (19) GB (11) 2 192 746 (13) A

(43) Application published 20 Jan 1988

(21) Application No 8716495

(22) Date of filing 14 Jul 1987

(30) Priority data

(31) 8617389

(32) 16 Jul 1986

(33) GB

(51) INT CL<sup>4</sup>

G10L 3/00

(52) Domestic classification (Edition J):

G4R 10A 10E 10F 11D 11E 11F 1F 3C 3E 3F 9B HA  
U1S 1727 G4R

(71) Applicant

British Telecommunications public limited company

(Incorporated in United Kingdom)

81 Newgate Street, London EC1A 7AJ

(72) Inventors

Ian Ross Cameron

Paul Christopher Millar

(74) Agent and/or Address for Service

B G W Lloyd,

Intellectual Property Unit, Room 1304, 151 Gower  
Street, London WC1E 6BA

(56) Documents cited

GB A 2170936

GB A 2098773

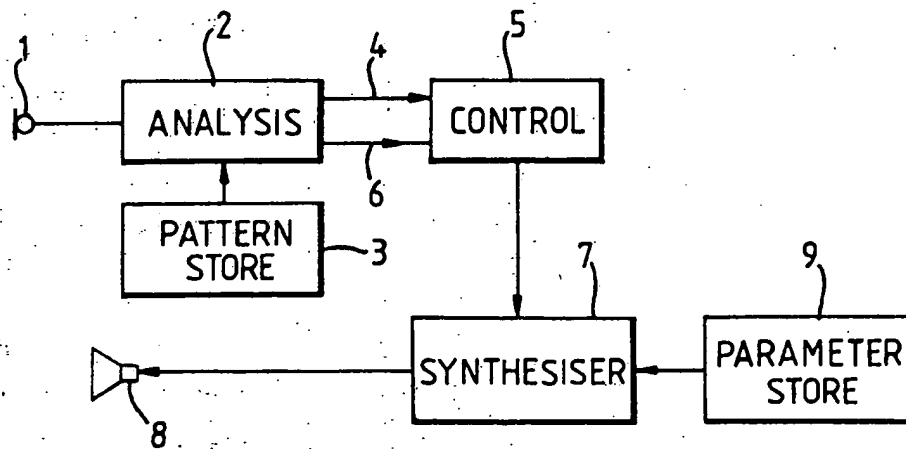
(58) Field of search

G4R

Selected US specifications from IPC sub-class G10L

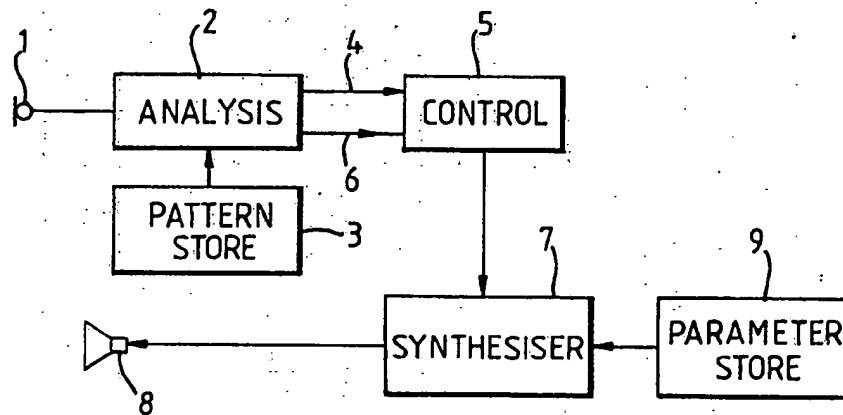
## (54) Speech recognition

(57) Words uttered by a user are compared 2, with stored words 3; the word giving the best "score" in the comparison is deemed to have been recognised. Where equal or similar scores occur the result is ambiguous and in that case a message is generated (eg by means of a speech synthesiser 7) containing a word for the user to confirm. If he does not, a second word may similarly be offered.



2192746

///



## SPECIFICATION

### Speech recognition

- 5 The present invention relates to speech recognition systems. In such systems words uttered are subjected to known pattern recognition techniques and if correspondence with a known word is found, suitable coded signals are generated identifying the word. Correspondence is generally determined by generating signals or "scores" indicating the degree of similarity with stored patterns corresponding to known words; the word having the best score is deemed to be the word uttered. This technique fails, however, if an ambiguous result is obtained (ie if two scores are obtained which are the same or differ only by a small amount). Normally in an interactive arrangement the remedy is for the recognition system to respond by presenting the user with a request to repeat the word in question.
- 10 However, this approach suffers from the disadvantage that there is a high probability of the ambiguity recurring; also it can be irksome for the user.
- 15 According to the present invention, therefore, there is provided a speech recognition apparatus comprising analysis means for receiving speech signals from a user, comparing each received word with stored representations of words to produce similarity signals indicating the degree of correspondence between them, and producing coded signals identifying recognised words, and
- 20 output means for presenting messages to the user, the analysis means being operable in the event that the similarity signals in respect of a first stored representation to which a received word most closely corresponds is equal to, or differs by less than a predetermined margin from, the similarity signal in respect of a second stored representation to:
- (a) generate via the output means a message including the word represented by the first stored representation;
- (b) await an indication from the user as to whether the word is correct;
- (c) upon receipt of a positive indication to generate the said coded signal.
- In the event that a negative indication is received from the user, the analysis means may generate a message requesting repetition of the word, or may
- 30 (i) generate via the output means a message including the word represented by the second stored representation and
- (ii) await an indication from the user as to whether the word is correct.
- The output means may be a visual display, or could be a speech synthesiser.
- The indication from the user may be input by means of switches or a keypad, but more preferably is by speaking appropriate words (eg "Yes" or "no") which may then be analysed by the analysis means.
- 35 One embodiment of the invention will now be described by way of example, with reference to the accompanying drawing which is a block diagram of a speech recognition apparatus.
- In the figure, speech from a user is received by a microphone 1 connected to a speech recogniser 2. The recogniser compares received words with the contents of a pattern store 3 which contains representations of a repertoire of words which it is desired to recognise.
- Any of a number of conventional recognition algorithms may be used and these will not therefore be discussed in detail. By way of example the "VOTAN" recogniser card produced by Votan Inc. for use with an IBM PC microcomputer might be employed.
- 45 The recogniser 2 compares a received word with each of the stored representations and produces for each a similarity signal or "score" which indicates the closeness of fit between the two. Normally the word whose stored representation has the best score is the one "recognised" and a corresponding coded signal is passed via line 4 to a control unit 5, which could for example be the aforementioned IBM computer, for onward transmission or to initiate further
- 50 action, according to the purpose of the system.
- If, however, two stored representations have the same or similar scores, the result is ambiguous and a signal indicating this is passed via line 6, along with codes for both words via line 4, to the control unit 5 which responds by generating a message back to the user via a speech synthesiser 7 and loudspeaker 8.
- 55 This message has the form of "Did you say X", where X is the word whose representation stored in the pattern store gave rise to the better score (or, if the two scores were identical, one of the two selected at random); and awaits a reply. The synthesiser is assumed to have a parameter store 9 to enable it to generate appropriate words.
- If the user replies "Yes" (or "No") this is recognised by the recogniser 2 and signalled to the control unit 5 which, in the event of a Yes proceeds as if X had been identified originally. In the event of a "No", a further message is issued via the synthesiser, viz "Did you say Y". Again the user response is analysed and if Y is confirmed, recognition is deemed complete. If the user again replies "No", the control unit then initiates generation of a request for repetition (although in principle of course the third choice could be offered).
- 60 By way of example, 1/26/05, EAST Version: 2.0.1.4
- 65 By way of example, 1/26/05, EAST Version: 2.0.1.4 might be in a telephone banking

service. Here the control unit would be programmed to generate questions to the user, via the speech synthesiser, and to respond by generating further questions to elicit the required information to assemble an instruction which may then be passed to a bank's staff or computer, for effecting a credit transfer, printing a statement, or the like.

- 5 A typical set of words representations of which might be included in the pattern store 3 might be 5

Services:

- "Statement" -Order full statement  
10 "Balance" -Give Balance 10  
"Mini-statement" -Last 4 transactions  
"Transfer" -Transfer money between accounts  
"Cheque-book" -Order new cheque-book  
"Help" -Request assistance from bank staff

- 15 15

Account Types:

- "Current" ) Bank accounts  
"Savings" )  
"One" ) Credit card accounts  
20 "Two" ) 20

Amounts:

- "Ten" )  
"Twenty" )  
25 "Thirty" ) Amount in pounds 25  
"Forty" )  
"Fifty" )  
"Full" -Make full payment

- 30 Cancel: 30  
"Stop" -Cancel service (may be used during speech output)

A typical user-machine dialogue might proceed as follows (after an entry procedure with appropriate identity numbers and/or passwords-possibly accompanied by speaker recognition techniques for added security):

- 35 35

- (1) System: "Which service do you require?"  
(2) User: "Transfer" (mispronounced)  
(3) System: "Did you say 'statement'"  
40 (4) User: "No" 40  
(5) System: "Did you say 'transfer'"  
(6) User: "Yes"  
(7) System: "From which account do you wish to transfer funds?"  
(8) User: "Savings"  
45 (9) System: "Which account do you wish to transfer funds to?" 45  
(10) User: "Current" (mispronounced)  
(11) System: "Did you say 'Current'"  
(12) User: "Yes"  
(13) System: "How much money (in pounds) do you wish to transfer from  
50 your savings account to your current account?" 50  
(14) User: "Ten" (mispronounced)  
(15) System: "Did you say 'twenty'"  
(16) User: "No"  
(17) System: "Did you say 'thirty'"  
55 (18) User: "No" 55  
(19) System: "How much money (in pounds) do you wish to transfer from  
your savings account to your current account?"  
(20) User: "Ten"  
(21) System: "Ten pounds will be transferred from your savings account  
60 to your current account. Do you require another service?" 60  
etc

Note the statements at lines 2, 10 and 14 where poor pronunciation, noise or the like has given rise to an ambiguity which has been resolved in two cases by offering to the user the words judged to be closest t1/26/05, EAST Version: 2.0.1.4

## CLAIMS

1. A speech recognition apparatus comprising analysis means for receiving speech signals from a user, comparing each received word with stored representations of words to produce similarity signals indicating the degree of correspondence between them, and producing coded signals identifying recognised words, and output means for presenting messages to the user, the analysis means being operable in the event that the similarity signals in respect of a first stored representation to which a received word most closely corresponds is equal to, or differs by less than a predetermined margin from, the similarity signal in respect of a second stored representation to
- 5
- 10 (a) generate via the output means a message including the word represented by the first stored representation;
- 10 (b) await an indication from the user as to whether the word is correct;
- (c) upon receipt of a positive indication to produce the said coded signal.
2. An apparatus according to claim 1 in which the analysis means is arranged, upon receipt of a negative indication to (i) generate via the output means a message including the word represented by the second stored representation and (ii) await an indication from the user as to whether the word is correct.
- 15
3. An apparatus according to claim 2 in which the output means is a speech synthesiser.
4. A speech recognition apparatus substantially as herein described with reference to the accompanying drawing.
- 20

Published 1988 at The Patent Office, State House, 66/71 High Holborn, London WC1R 4TP. Further copies may be obtained from The Patent Office, Sales Branch, St Mary Cray, Orpington, Kent BR5 3RD. Printed by Burgess & Son (Abingdon) Ltd. Con. 1/87.